CHAIRMAN MEHTA: Dr. Hooley, any questions?

DR. HOOLEY: I have a question in line with Dr. Conant's. The presentation emphasized mammographically-detected masses, but I'm unclear if the study included masses that were detected only on clinical breast exam.

DR. PARISKY: Yes, it did.

DR. CALLAHAN: The original study protocol was intended to analyze both patients that were enrolled based on clinical exam alone, say, you know, palpable lesions that weren't identified, that weren't visible mammographically. That was the original study protocol.

But when it became apparent that really the utilization of this device requires the evaluator, the physician, to localize on the IR image, and because we had this independent panel of reviewers that did not have access to the patient, if it was a palpable lesion that wasn't visible mammographically, then they couldn't perform a localization and assessment.

So, in the end, our efficacy claim would be that it's a mammographically-apparent mass that can be localized. Dr. Hughes may comment on, and maybe perhaps you want to defer this discussion until later, but Dr. Hughes can comment on the protocol at Mass. General, where I believe they are enrolling patients with palpable lesions, because the enrolling physician can go ahead and localize based on the palpable lesion location that they are aware of without having mammographic evidence. But our clinical trial could not assure that, because our evaluators were blind and independent.

DR. HOOLEY: I have one more question.

CHAIRMAN MEHTA: Go ahead.

DR. HOOLEY: The cost that you cited for the biopsies of $3,000 per biopsy, I thought that was in a very high range. Yet, you used, your range for your procedure, you used like $220 or something like that, giving a medium range of between $150 and $300, but $3,000 for a breast biopsy seems excessive.

DR. PARISKY: In defense of that, that number is taken from a compilation of both core biopsy

and surgical biopsy. Since the national average is approximately 50/50, while it may not be in academic centers and certain centers of excellence, of a number of mammographically-apparent lesions still undergo needle localization, and we'll probably continue to do so, which has a much higher cost. That's how that figure was arrived at.

And it's not particularly different, accounting for inflation and rise in health care costs, from the numbers quoted by Jim Brenner and Ed Sickles from their article about seven years ago, which I believe was $1,900 or $1,700. So taking that into account, I think $3,000 is a fair number.

DR. CONANT: How about the cost of the followup for those IR-negative lesions, the six-month followups and the cost of those women who were imaged and uninterpretable?

DR. PARISKY: Well, the uninterpretable I will address that first. There was a learning curve for both the machine and I think for the technologists in terms of uninterpretable. Steps have been taken; again, as a practicing clinician, I want to minimize

the number of repeats, and so forth. So I believe that that has been mitigated.

In terms of followup, patients who you would, in the absence of this technology, would follow up as well, based on giving them a BIRADS 3, we didn't calculate that score, and that would be based on physician preference.

So you are talking about additional diagnostic mammogram or --

DR. CONANT: Would you recommend this as a followup as well?

DR. PARISKY: This would have a followup as well and the appropriate clinical judgment of the clinician. So there may be an increased cost, a slight increase in cost, in followup, but we didn't subtract out or we didn't calculate, if you conventionally sent patients to BIRADS 3, what the costs are with that.

DR. CONANT: Sure.

CHAIRMAN MEHTA: Go ahead, Geoff.

DR. IBBOTT: I have a couple of technical questions.

DR. PARISKY: Oh, good. I'll sit down.

(Laughter.)

DR. IBBOTT: The description of the use of the images focused on the central en face image, outlining the breast and the region of interest. How are the peripheral six images used or are they used?

MR. SATTERTHWAITE: You'll note from those images that there are anatomical features that are apparent. In the training that CTI does of both the technologist and the physician, we train them on how to localize or identify the breast by looking at outlines of the breast in the side mirrors and drawing lines across the en face image in order to clearly identify exactly where breast tissue is. So those are used primarily to identify nipple location and to outline the breast.

DR. IBBOTT: I see. What sort of calibration and quality assurance procedures are required? If this device were to be approved, how would those procedures be implemented in a routine setting?

MR. SATTERTHWAITE: The device that will

be marketed has a calibration that takes place before every imaging session. There's a black body device in there that allows us to calibrate each time.

There is regular maintenance on the camera that is associated just with the camera. Every year there's things that happen with the camera. There are other things that we do to assure quality, like check the mirrors and a number of things that are in that manual that has been provided the FDA.

DR. IBBOTT: And the region of interest was expanded by the software to something approximately one-twelfth of the area, I guess, of the breast tissue. How was that one-twelfth value selected? Because it seems like choosing a smaller area that's limited to the mass would make the device more sensitive.

MR. SATTERTHWAITE: I'll ask Dr. Rust to come and back me up here on that one. I mean, obviously, there was some work that we did, but we recognize the fact that to localize between modalities there's some variance just because of the presentation of breast tissue in those various modalities. So we

saw the need to improve that localization by looking for a stronger IR signal in the vicinity.

So the actual selection of the one-twelfth, I will let Dr. Rust address. I believe that was your primary question.

DR. RUST: Maybe I will start by clarifying how that procedure works. The area that is one-twelfth of the breast region is the size of the search region for looking for a point of both higher IR signal and higher contrast with surrounding pixels.

Once that location is determined, what is actually used for determination of a negative or a positive outcome is actually a circle of radius five pixels, much smaller than that one-twelfth region. So the one-twelfth area is a search region. The decision is made on a much smaller focused region of interest.

DR. IBBOTT: Thank you.

CHAIRMAN MEHTA: Dr. Rust, while you're up there, I did have one question for you.

DR. RUST: Sure.

CHAIRMAN MEHTA: You were very clear in outlining to us that the mass subset was prospectively

defined and the planned assessment was included in the protocol.

DR. RUST: Yes.

CHAIRMAN MEHTA: I do have a question that goes to the general concept of prospectively-planned statistical analysis and assessment in the protocol and have several subsections to the question.

Did the protocol state that this assessment for masses would be done after the biopsy results were available or before? Because it appears to me that this assessment was done after the biopsy results became available.

Secondly, did the protocol prospectively allow for an expansion in patient pool from 600 to 2,400, a fourfold increase? Was that prospectively defined in the protocol? Did the protocol prospectively allow the inclusion of 275 patients that would not be evaluated and would be unblinded at a later date to be evaluated?

These are sort of confusing issues and aspects in the conduct of the protocol, unclear whether this was prospectively planned for and, if so,

how did you convince IRBs about the justification of such a prospective plan?

DR. RUST: To your first point, the protocol stated that performance analysis would be done by lesion type, which implicitly requires that it be done after pathology is unblinded. Otherwise, it is impossible to look at lesion type --

CHAIRMAN MEHTA: I don't think a lesion typed labeled "mass" is a pathologic diagnosis.

DR. RUST: No, it's not, but what are you analyzing by lesion type? You are analyzing performance results, and there are no performance results to analyze prior to unblinding of the pathology. So implicitly I believe that those analyses were prospectively planned to be done after unblinding of pathology. Otherwise, there is nothing to analyze.

Your second question was the expansion of the --

CHAIRMAN MEHTA: Six hundred to 2,400.

DR. RUST: I do not believe that the expansion was prospectively planned for in the

protocol.

DR. CALLAHAN: The study protocols were submitted independently to each IRB, and each one stated the 600 enrollment definition. So each protocol was independently under the each institution IRB with a sample size of 600.

It wasn't clearly stated, a total enrollment goal, in the protocol, I'll say that.

CHAIRMAN MEHTA: I'm sorry, does that mean 600 patients from each site?

DR. CALLAHAN: Each site was independently, yes, provided the study protocol with an enrollment goal of 600 patients.

CHAIRMAN MEHTA: And you had six sites. That would require --

DR. CALLAHAN: Well, we had actually five, six enrolling centers. Dr. Parisky's site at USC involved LA County initially and then Norris Cancer Center. So it would have been 600 at that one site, with two enrolling centers, and then our four other sites.

CHAIRMAN MEHTA: So that would have

required 3,000 patients, 600 per five, per site, times five. Was the protocol prospectively designed to stop at 2,400?

DR. CALLAHAN: I cannot answer that, as I was not there at the initiation of these study protocols that began 1998. I don't have that information. I don't know that any member of our team has that information.

There was no, as far as I know, there was no prospectively-defined written document saying that the total enrollment goal would be a specific number.

CHAIRMAN MEHTA: Will you, at a later point in the meeting, be able to provide us site-specific enrollment data?

DR. CALLAHAN: We have that on one of our slides --

CHAIRMAN MEHTA: Okay.

DR. CALLAHAN: -- if you want to go back. I can state that they range from the lowest enrolling site was 170, and our highest enrolling site was a little over 800.

DR. GENANT: I have a question.

CHAIRMAN MEHTA: Go ahead.

DR. RUST: There was a third part to your question.

CHAIRMAN MEHTA: Yes.

DR. RUST: Would you like me to address that?

CHAIRMAN MEHTA: If you could.

DR. RUST: Okay. I can simply tell you what happened in terms of how the 275 patients were actually identified. There was a schedule put in place to complete analyses, to go into Module 5 for submission to the FDA. In order to support that schedule, a freeze was placed on the clinical database in terms of patients enrolled before a particular date. Then that frozen database was taken forward through all the analysis procedures, results analyzed, put into the original Module 5 submission, and submitted.

Centers continued, three of the centers continued to enroll patients after that database freeze, and that was the source of the additional 275 patients that ultimately were analyzed in the

confirmatory study.

CHAIRMAN MEHTA: Okay. Go ahead.

DR. GENANT: Yes, I have questions with regard to the reproducibility, No. 1, of the machine itself. Do you have data that would indicate that, if a woman was measured on two occasions, that there would be comparability of results?

And, secondly, you have pooled the readers' results, and I wonder, do you have information on the individual readers and how they related to each other, what the reproducibility was, and how would that extrapolate to clinical practice, where one would have a single reader?

MR. SATTERTHWAITE: I'll address your first question, which is the reproducibility. We have not imaged a single patient in two different situations. We have not addressed that.

DR. RUST: In terms of pooling the information from multiple evaluators, let me describe exactly how we do that. Our analyses were all lesion-based, and we took the three evaluations of a particular lesion and included all three evaluations

in the analysis, giving each one a weight of one-third in the analysis, taking credit for only one lesion with that analysis.

So we did not pool the three evaluators into a single positive/negative score which was then analyzed. We did include each individual evaluation, giving it a weight which caused each lesion to get a total weight of one in the analysis. So I wanted to clarify that.

So, to answer your question, do we have data on the individual evaluators' scores, yes, we do, and it was incorporated into the analysis in the fashion that I just described.

DR. GENANT: Can you share with us what that inter-reader variability was?

DR. RUST: In fact, we did an analysis to answer a question that the FDA posed to us on inter-reader variability in the IOS score that leads to a positive/negative test result. I'm going to have to back up here and describe what the IOS score is.

It is an index on a scale of zero to a hundred, and small values are associated with less-

suspicious lesion, large values with more-suspicious lesions. We compare that to a threshold to get the positive and negative test result.

The result of our analysis on inter-reader variability was that the inter-reader standard deviation was approximately four units, which in this context appeared to be rather small, both to us and the FDA, I believe.

Does that address your question?

DR. CONANT: May I ask a question along that line? That's for the IR part. I am very interested, I think I mentioned before, about the inter-reader variability, about what brought the patient to the category of biopsy necessary. Do you have that information?

DR. RUST: Okay, I don't believe I can address that question.

DR. CONANT: Because there's at least a 30 percent variability, unfortunately, I think, in most practices.

Another question similar to that is this threshold idea. I read about the threshold and how it

was chosen. I am very curious about the distribution of the IR readings about threshold and how they correlated with the LOS, the level of suspicion, in terms of the radiologist.

It sounds like a great threshold. I mean, it only missed one. But I am wondering how close things really were in there and whether they were all very borderline, and there's a whole group of gray zone cases --

DR. RUST: Right, we --

DR. CONANT: -- because those are very difficult clinically.

DR. RUST: We did, in fact, do an analysis of sensitive -- "sensitive" may be a bad word -- sensitivity of the performance parameters to the threshold selection and basically summarized that in the form of a plot showing the changes in sensitivity and specificity near that threshold.

Does that address your question at all?

DR. CONANT: Well, actually, I would like to see the raw data and how it fell, not just the sensitivity. I have questions with the sensitivity

because there are so many excluded patients. So I would rather look at the raw data.

I mean, it's an interesting -- I was just trying to crunch the numbers here, but Dr. Parisky said at the beginning about 50 percent of the cases that go to biopsy are masses. Your population, I think, if I did the numbers right, was around 20 percent. So sensitivity I am not sure reflects the population, but actually raw data around that threshold level --

DR. RUST: I just want to get a clarification on your question. You think that only 20 percent of our study population presented with masses?

DR. CONANT: No, but it looked like, when you came down to that -- maybe I did my numbers wrong. Maybe we can talk about that later, but it's not of the overall, but when you take out all those exclusions, when you crunch down to the numbers that then actually are eligible and then get a -- it's on page --

DR. RUST: Actually, I believe that --

DR. CONANT: -- page 12.

DR. RUST: I believe that that ratio in our study, and correct me if I'm wrong, was 412 out of 875.

DR. TOLEDANO: What would that have looked like if you excluded the masses up at the top? So you excluded for those; you excluded for that. And I know that this is the way that this study actually happened, but I think what Dr. Conant is interested in, and what I'm interested in, is, what if you excluded, what if you subdivided masses from non-masses up at the top?

DR. RUST: We do have the data to do such an analysis. I can't guess at what the answer would be.

DR. CONANT: But then to look at the raw numbers of the IR without choosing -- I mean, I'm sure you have a very good reason for your threshold, but it's just, you know, I would love to know.

DR. PARISKY: You're interested in the IOS related to malignancy, malignancy distribution or --

DR. CONANT: No, no. I'm really

interested in -- that's one question, yes, but --

DR. TOLEDANO: Aren't you asking for the distribution of IOS in the malignants --

DR. CONANT: Yes.

DR. TOLEDANO: -- and the distribution of IOS in the benign?

DR. CONANT: Yes, yes.

DR. RUST: And that data could certainly be made available, yes.

DR. PARISKY: And we have looked at it.

DR. CONANT: Great.

CHAIRMAN MEHTA: I think Prabhakar had a question.

DR. TRIPURANENI: In the confirmatory study you had 275 patients, out of which you had 78 masses. The 275 patients, is it before excluding certain groups of patients or after?

DR. RUST: It is before. That is the total number of patients before any cases were withdrawn for the various reasons that Dr. Callahan indicated would cause patients to be withdrawn.

DR. TRIPURANENI: So if you take the 78

masses out of that 275 patients, the percent of masses in those patients comes to approximately 30 percent or so. The same number into the original study comes to approximately 432 patients. Is it the equivalent number within the confirmatory study with the original study?

DR. RUST: I'm not clear on your question. I'm sorry, I didn't follow all --

DR. TRIPURANENI: Two hundred and seventy-five patients now.

DR. RUST: Yes.

DR. TRIPURANENI: What is the counterpart number in the original study? Is it 432 or is it 840?

DR. RUST: Eight hundred and seventy-five is after exclusion of stuff.

DR. TOLEDANO: I could do it.

(Laughter.)

DR. RUST: It's on Dr. Callahan's slide. I believe it is where we say 1,660 masses and --

DR. TOLEDANO: Okay, do you want me to do it? Okay, 2,406 patients goes down to 1,432 in the FDA dataset, when you exclude all the unvaulted. Of

the 1,432, you have approximately 800 who were in the original FDA evaluable dataset, of which 432 had masses. In the 275 you had 171 who were evaluable, of which 69 had masses. Did I get that right?

DR. RUST: I believe the comparable number is 1,432.

DR. TRIPURANENI: Then in that case, the percent of masses from the original study to the confirmatory study actually is almost double the number. Fourteen percent of the original values actually had the masses whereas in the confirmatory study it is almost 29 percent. Were there significant enrollment differences between the original study to the confirmatory study?

DR. RUST: Okay, I'm thinking the comparable ratio here is 412 to 1,432 versus 78 to 275, and are those really that different, I guess it the question I'm asking.

DR. CONANT: I think they both are in the 20 percent range, low twenties.

DR. RUST: I'm calculating on my feet here, but they don't seem all that different.

DR. CONANT: I think they're similar --

DR. RUST: Uh-huh.

DR. CONANT: -- but the question is, how does that deviate from what goes to biopsy in reality, which is about 50 percent masses? I'm not sure it's that high, but I don't have that number in front of me. I think it's lower.

DR. CALLAHAN: I think what we need to do to do this calculation is look at, like Dr. Toledano suggested, look at the number of masses in each of these groups prior to any exclusion to see what the percentages were and to see if there's differences, and that's something we can certainly do over the break.

CHAIRMAN MEHTA: I think we'll go to it at this time in order to stay on time. Let's go ahead and take a five- to ten-minute bathroom break. We'll reconvene at 11:00, so that the FDA can do its presentations. We will have another opportunity for questions later on in the afternoon.

(Whereupon, the foregoing matter went off the record at 10:54 a.m. and went back on the record

at 11:05 a.m.)

CHAIRMAN MEHTA: If could start having everyone in their seats, the first speaker will be Jack Monahan, the leader for the PMA.

MR. MONAHAN: Good morning. I would like to take this opportunity to thank the Chair and the other members of the Panel for taking time from their busy schedules to help us on the deliberations associated with this PMA.

As you know, we have had the sponsor present, and I would like to just give a few introductory remarks prior to moving into the clinical and statistical discussion.

This particular PMA is what we refer to as a modular submission, and we received the first module back in 1999. I was the primary reviewer for that module, which contained the preliminary information related to the device.

Module 2 was subsequently submitted, and that contained primarily the software material, and that was reviewed by Joseph Jorgens in our Office of Science and Technology.

Module 3 consisted of all of the manufacturing information, and that was reviewed in the Office of Compliance by Xuan Vo.

The fourth module was the engineering material associated with the product, and that was reviewed by Jim Seiler, who is in the Division.

The PMA was submitted and contained principally the clinical data, the protocol, and the labeling material for this product. This is what we are dealing with today.

I am the lead reviewer on the PMA. The clinical reviewer is Dr. William Sacks, who you will be hearing from in a moment. The statistical reviewer is Harry Bushar, and we also had a bioresearch monitoring review of this submission, and that was done by Kevin Hopson.

What I would like to do is very briefly go over the proposed indication for use. I have abbreviated a little bit on the screen, but you will note that I have highlighted some of the words in there because I believe that these are important aspects of this proposed indication for use.

The CTI BCS 2100 is a dynamic, computerized, infrared-based imaging, image acquisition and analysis system. It is intended for use as an adjunct to mammography, to safely avoid biopsies of benign breast masses that would otherwise have gone to biopsy.

Physicians should not base a decision for patient care solely on the results of testing with this device, but rather on results of this test in combination with all other findings and risk factors associated with a specific patient.

The CTI BCS 2100 provides additional information to guide a breast biopsy recommendation. Because demonstration of device effectiveness was limited to breast lesions that included mass as a lesion descriptor, use of the CTI BCS 2100 should be limited to the evaluation of breast lesions that include mass as a lesion descriptor. The presence of another lesion descriptor does not contraindicate use of the CTI BCS 2100 if the lesion is also described as a mass.

It is recommended that the appropriate

recommendation for care of all patients receiving a negative IR test result be similar to the recommendation for care of a mass that is assigned a mammographic category of 3 or a BIRADS 3. That is short-interval followup is recommended in order to establish the stability of the findings.

This is the proposed indication for use, and I would now like to turn this over to Dr. William Sacks to discuss the clinical study.

DR. SACKS: Just for those of you who don't know me, I'm a radiologist and used to be a physicist. So I have some familiarity with numbers as well.

I want to stress a number of aspects of the device that I will enlarge on as I go on. First of all, this is a new type of thermographic device. Secondly, it's an adjunct to mammography. Thirdly, it renders a positive or negative result, as you have seen, and that is based, as the company has explained, on an index-of-suspicion score.

It is intended for women on their way to biopsy only. It is, furthermore, intended for women

on their way to biopsy who have mammographic masses only, and the intended use is to save biopsies of lesions that turn out to be benign.

The points I am going to cover are what the BCS is and is not intended to do; how the device does it. Then I'm going to have Dr. Bushar give the clinical trial results, and I will come back and make some assessment of those results and, finally, a few labeling issues that we would like the Panel to consider.

Before I start on what the BCS is and is not intended to do, I want to make a clear distinction in the minds of the Panel between a device and its intended use. It is very important to keep that in mind as we go on. One and the same device can have a number of different intended uses, and, indeed, for any given intended use, there may be one or more devices that will satisfy that use.

We will be talking predominantly about the clinical trial. A clinical trial is always designed based to demonstrate that the particular chosen intended use of the device is safe and effective. So

it will always be an underlying issue here that we are talking about the company's intended use for this device.

Now in case any of you come here with any baggage or prejudice from the past about breast thermography, I want to make a clean break with that. Historically, it has not had the sensitivity and specificity to either replace screening mammography or to be a complementary screening test; that is, it hasn't had the sensitivity or specificity to be a screening test.

However, the BCS is a new type, as I said, of thermographic device, and it is new in two ways. One, it uses a new application of technology which lies predominantly in the cooling of the breast with the fan, and that enlarges the temperature contrast between malignant tissue and benign tissue, it is thought. The reason for that is that the benign tissue will cool, whereas the malignant is fed by angiogenesis and has a higher metabolic rate, will not cool as fast.

So that if you were to track the time

course, as this device does, over the cooling, the contrast between malignant and benign tissue would be enhanced. So that is one aspect that is new.

The second one is that it targets a different group of women from that which conventional thermography in the eighties tried to target. In the eighties the attempt was to make this a screening device that, hopefully, would replace mammography or at least work alongside it for all women screened. This device, however, targets, as you have heard, and as I myself have mentioned, a subgroup of screened women.

I've already mentioned that, that the cooling is the issue here, and the different group of women is the ones whose screening tests, mammography and/or palpation, along with other factors, indicate a need for biopsy.

So it is not intended as a screening device -- that's very important -- and it's neither, therefore, a replacement nor a complement to screening mammography, but rather as an adjunct, and, in particular, an adjunct to mammography, not to clinical

palpation, as was hoped at the time of the original protocol, but to mammography, and, indeed, only for women on their way to biopsy and only for those with mammographic masses.

Let me say a few words about the difference between a complementary test and an adjunctive test. These are somewhat confusing concepts, and adjunctive is itself probably the most confusing.

Let me say something about complementary tests to begin with. A test that's complementary to a screening test is used on all persons screened; that is, it is itself a screening test, and, therefore, its results may by themselves determine the next step in clinical management.

Complementary screening tests, therefore, are on equal footing with each other. One easy example is screening mammography and clinical examination. Women over 40 get annually, should get annually, a clinical palpation as well as screening mammography. If either one of these shows need for a biopsy, such as a palpable mass, even if it's

invisible on the mammogram, then the clinical examination will be the thing that will decide the woman's clinical management.

If, on the other hand, there is nothing to palpate, but the mammogram shows a suspicious finding, the woman will still go on and get further workup. So these two exams are complementary to each other because either one by itself can determine the next step.

Adjuncts, on the other hand, are subordinate to the index screening test; that is, the screening test to which they are adjuncts. They can be subordinate in one of two ways or both.

They are either not used on all the persons screened, and I will come back to the examples in a second, or if they are used on all the persons screened, their results do not by themselves determine the next step in clinical management. Let me give you examples of each of these.

The BCS itself is the first type. It is not used on all persons screened. However, on those on whom it is used its results do generally by

themselves, or will, or it is intended that, its results will determine the next step in clinical management; namely, whether or not the woman goes on to biopsy or not.

In a sense, it is inherent in the nature of the device, which is a black box that pops out a number, you can't make a judgment, the company has stressed. You have no -- it is not a visual issue of the image itself; the device gives you a number. So insofar as it does, you are forced to listen to the device.

Now you are not forced to do what the device tells you, but if you do, it is not a -- let me put it this way: If you have a woman that you really think needs a biopsy and you subject her to this test, and this test gives you a negative result and you decide to send her to biopsy anyway, my suggestion is, don't do the test. There's no point in having done it. You could have foreseen that ahead of time.

Another example of an adjunct of this type is the ultrasound of solid breast masses, as it is being done by a growing number of people, Stavros and

others, and so on. It is used only on a subgroup of those who go through mammography or palpation, and it may by itself determine whether the woman needs a biopsy or not, if you happen to use ultrasound in that fashion for solid masses. Extra-mammographic views are another example, and so on, and even biopsy itself.

Now the other type, the results by themselves don't determine the next step in clinical management, even if they are used on all the women screened -- and a perfect example of that is a mammography computer-assisted diagnostic system. It is used on everybody, but it is the radiologist who decides, after it points out places, "Have you looked here, here, and here," whether or not to do something about that.

So those two types of subordination, either one of those or both will throw a device into an adjunctive status.

The intended use -- and I stress it again -- the intended use, as currently intended, of the BCS is to confirm the need for biopsy or change a woman's

clinical management. If it is changed, it is to be changed from biopsy only to short-term followup, not to come back in a year for the next screen.

Thereby, it can only decrease the number of biopsies, and in statistical language that means it can only increase the specificity. It cannot increase the detection of cancers as it is currently intended to be used and as the trial was conducted. That is, it cannot increase sensitivity.

An advantage of the particular selection of target population -- that is, only women on their way to biopsy -- is that device false positives, and I define a device false positive as a woman who has a mass that is benign but gets a BCS-positive result. It's that simple. Such device false positives have no impact on clinical management. After all, these are women who were on their way to biopsy anyway, and if you get this positive result, you will simply go on and do what was recommended in the first place. Therefore, there's no impact upon clinical management.

We would like the Panel, during the discussion this afternoon, to consider an issue as to

whether they have any concern over the potential psychological impact of a positive mammogram followed by a false positive BCS result, that is, on a woman who does not, in fact, have cancer. Let me just say a word about that.

If I were to get a mammogram based on which a recommendation that I get a biopsy was made, my main fear would not be of a biopsy procedure; it would be that I had cancer. Now I'm offered a test that says we can do one other test that may obviate the need for a biopsy, and if I get a positive result from that test, now I'm a little more convinced that I must have cancer, even though -- and it can be explained to women, and this is what we want you to discuss, whether labeling or anything like that needs to be addressed -- is your chance of having cancer zoomed from about 20 to 23 percent. I mean, you're still overwhelmingly not likely to have cancer, even though both tests were positive, but this is a subject that one of our questions will be designed to ask you to discuss.

Is the BCS an alternative to biopsy? In

80 to 85 percent of women who obtain the test, it will end up being in addition to biopsy. Only 15 to 20 percent of such women will end up not getting a biopsy in addition, at least not immediately.

How does the BCS do this? This is somewhat repetitious, but it is good to hear a little redundancy when so much information is being thrown at you.

It calculates an index-of-suspicion score for the region of interest selected by the radiologist, and the radiologist bases that selection on the mammographic location of the mass. The device then compares whatever that number is, which ranges from zero to a hundred, with the determined threshold that was determined by the company during their training set of the first 700 -- a slight detail there, but Dr. Bushar will talk about that -- but, roughly, the first 700 out of the 2,407 patients, did some training and picked the threshold of 20.59 so as to keep a very high sensitivity.

If the IOS score for this woman falls below that threshold, the device will read negative

results, and that means change her path to biopsy to short-term followup. That is very similar to the BIRADS 3 category. If the IOS score is at or above the threshold, the device output will read positive, and that means just continue with the plan to biopsy.

There is a side effect, and that is that some cancer diagnoses may be delayed, and we can talk about that.

Now I would like to have Dr. Bushar give you some of the statistics here.

DR. BUSHAR: Thank you very much, Bill.

Good morning. My name is Harry Bushar. I'm the statistician who reviewed this PMA on the computerized thermal imaging Breast Cancer System 2100. I will be doing the statistical presentation. An outline of what I will be presenting is I want to discuss the clinical study protocol, including objective design, population, demographics, and evaluation, both effectiveness and safety, and then get into the actual PMA clinical study and what was done there in terms of effectiveness and safety.

And, finally, continue to move on to

Amendment 4, which again brings out the effectiveness in the clinical study, and Amendment 5. Both of these amendments were caused by letters that the FDA sent, deficiency letters, to the company. These are their responses to those. And, finally, Amendment 7, which gives an adjustment to the effectiveness, and I want to make some statistical conclusions.

I would like to make it clear that what I did is I reviewed the sponsor's analysis. I did not actually analyze the sponsor's data. So what I will be showing you are the actual results of the sponsor.

In the clinical study protocol, the study objective was to determine if the CTI system, when used in conjunction with clinical examination and/or diagnostic mammography, increases the ability of physicians to differentiate benign from malignant or suspicious breast abnormalities.

Now what had to be done at a later date was to drop the clinical examination alone because they found they could not focus the BCS unless they had the actual mammography results.

Continuing with the clinical study

protocol, the study design, this is a prospective study. There's a blinding to histology, which meant that the actual histology results were not made available until the data was actually analyzed.

It is a multi-center study. There were actually six physical sites. The design of the study was intended to compare the level of suspicion -- that's LOS -- which is a number, 0, 1, 2, 3, 4, 5, of malignancy of suspicious breast lesions for clinical examination or diagnostic mammography before the BCS used, through the combination score LOS plus the BCS index of suspicion, IOS, which is a score ranging from zero to one hundred and measured in hundredths. So it is almost a continuous score, in contrast to the very discrete score of the LOS. This is, again, of malignancy of suspicious breast lesions.

Biopsy was used as the gold standard for pathology. One thing that had to be changed there was, again, clinical examination had to be dropped because the diagnostic mammography was required for the BCS to be used.

In the study population, the original

study population was 600 patients with biopsy. The actual study population was 2,407 patients with biopsy.

The demographics: gender, there were actually 15 males; ethnicity, mostly Caucasian, a good deal of African-Americans, some Latinos, and others; age, most people were 40 to 60 with a lot being greater than 60.

The primary effectiveness was on the overall population. The evaluation was to be the area under the ROC curve, the AUC, to compare results of diagnostic LOS score without BCS and the diagnostic mammography with BCS. That is a combined score, LOS plus IOS.

They also mentioned that they would look at sensitivity and specificity, and the CTI system will be considered effective if its performance in conjunction with diagnostic mammography and/or clinical examination is clinically better than mammography and/or clinical examination alone.

ROC is the receiver operating characteristic.

There were secondary effectiveness mentioned on subpopulations. They didn't say what they were going to do with this, but they did say they would look at the mammography lesion type by three categories: calcifications, masses, and distortions. These lesion types are not mutually-exclusive categories since a lesion can be more than just one.

The mammographic lesion size was measured in three categories broken into a half centimeter and one centimeter, and the mammographic lesion depth was simply stated "as available in the protocol." It wasn't specified.

Safety evaluation was by occurrence of adverse events. In the PMA clinical study population, the sponsor acquired BCS images from 2,407 patients at six physical U.S. clinical sites from December of 1996 through April of 2001. The sponsor actually analyzed only those patients with both mammography, not those with just clinical examination, and biopsy within 60 days.

Bob, could you turn on the slide? This will show a flowchart of the sponsor's clinical study,

starting at 2,407 and then fanning off in various directions, depending on what part of the study I'm talking about. Hopefully, that will help. I don't know whether that helps or not.

(Laughter.)

Look at your hard copy. It's a lot easier to read than what's on the screen, but if you want to glance up there, it's there.

The PMA training clinical study consisted of 700 patients. This consisted of the first 220 patients plus an additional 480 patients which were randomly selected from among the next 1,912 patients. These were used by the sponsor to set the following BCS IOS, index of suspicion, cutoff, and that is 20.59, which implies a recommendation for biopsy of a given lesion, or less than 20.59, which implies a recommendation for short-interval followup of a given lesion.

There were 1,432 patients enrolled from December 1996 through October of 2000, and these were initially available to test the effectiveness of the BCS, both in the original PMA and in Amendment 4. Now

what's missing here are the 275 patients which were in the pipeline at the time the data was frozen and analyzed.

Out of these 1,432 patients, there were 769 patients with 187 malignant and 688 benign lesions that were actually included in the effectiveness evaluation. This is true both for the original PMA and for Amendment 4. The population didn't change from one submission to the other.

Note that each patient had from one to four lesions, and the sponsor assumes that lesions within patients are independent in their analyses.

The PMA clinical study results: The primary effectiveness was by ROC, area under the curve. I'm just going to mention that the last two analyses the sponsor did, because I think they make a point, the sponsor found, after excluding calcifications alone, a statistically-significant greater area for the combined score IOS plus LOS1. By "LOS1," I mean that this is an LOS score where the unknowns and zeroes were eliminated.

Then for mammography, LOS score, which

score is 1, 2, 3, 4, alone had a significance level of .05. Now what they then found after, again, excluding the calcifications, but now expanding the mammography LOS1 score to add an additional two intermediate categories, 3.5 and 3.75, which was obtained by rereading the 4s, no statistically-significant difference in the area. This is referred to as LOS2 because now it includes the numbers 1, 2, 3, 3.5, 3.75, and 4.

Now I have done some plots here. What I did is I just crudely reproduced what the sponsor has in the PMA submission. You can see there that the purple line indicates the combined score, the IOS plus LOS1, which is essentially the IOS score just bumped up or down, depending on what the LOS was. You can see that's almost a continuous curve, and you can see what happens when you use the -- you can see the purple curve is almost continuous. This represents the IOS score bumped up or down by the LOS.

Now when we go to LOS, we only have a few numbers. We have 1, 2, 3, 4. You can see there's a point here, and there's, of course, an obvious point

down there. What has happened is that those two points are just connected with a straight line, which makes the area under the LOS curve smaller, statistically-significantly smaller, than the area under the combined curve.

Now I want to show that that is an artifact because, if we go to the next curve, now I have added in two more points. I have added in just two, 3.5 and 3.75, and you can see now the curve, the LOS curve, moves up. In fact, it even intertwines with the combined score.

Here, although physically the area under the IOS-plus-LOS curve is greater, the statistical significance is lost. I think this shows the problem of trying to compare a continuous curve with a discrete curve. There's a definite bias against the discrete curve.

Now continuing with the PMA clinical study results for safety, the following four adverse events occurred out of 2,407 subjects. This is from December 1996 all the way to the end, April 2001.

There were two mild, possibly-related,

resolved adverse events with both associated with patient discomfort during positioning, and then there was one serious and one mild, not-likely-related, resolved adverse event for hospitalization for treatment of a preexisting metabolic disorder and dizziness when sitting up after thermal imaging, respectively.

Now with Amendment 4, the sponsor essentially dropped their ROC analysis and just focused on sensitivity and specificity. This was mentioned in the PMA, but no detail was given. Here we have a little bit more detail on what the sensitivity and specificity look like using the 20.59 cutoff.

You see overall, with 187 malignant, to determine the sensitivity, and 688 benign, to determine the specificity, we get a sensitivity around 97 percent with a 95 percent confidence interval from 94 to 99 percent and a specificity of 14 percent with a confidence interval from 12 to 17 percent.

Now this overall result was rejected in terms of looking at the various lesion categories.

The numbers given here are a little bit strange because they don't add up, and the reason they don't add up is some calcifications also were masses, and whatnot. So what's being done here is any calcification is shown at the top, any mass in the middle, and any distortion at the bottom.

You can see from this that they got perfect results for both masses and distortions in terms of sensitivity. That is 100 percent. The sponsor interpreted these effectiveness clinical results by lesion type to specifically exclude that didn't come out to be 100 percent; namely, the calcifications, where it was only 95 percent sensitivity, which they claimed was not acceptable. And you can see the confidence intervals associated with those estimates.

The Amendment 4 interpretation: The sponsor's initial rejection of the overall effectiveness, followed by the sponsor's differential findings among the three lesion-type subpopulation, clearly indicates exploration, which does require confirmation and must be based on new data.

That came in Amendment 5, in which they did the so-called post-PMA or PPMA population. These were the 275 additional patients that had been in the pipeline at the time of the original submission and now were analyzed for the first time.

The gender is almost the same as before, which is one male; ethnicity, mostly Caucasian, some African-Americans, and a few others; age, again, mostly 40 to 60 with some large number greater than 60. So very similar to the original population.

These 275 additional patients were enrolled from November now of 2000 through April of 2001 at just three of the six original U.S. clinical sites which were initially available for confirmation of the effectiveness of the BCS in Amendment 5. Out of these 275 additional patients, there were 173 patients after exclusions with 43 malignant plus 159 benign lesions that were actually included in the Amendment 5 effectiveness evaluation.

Similar to the original, these patients had from one to three lesions. Again, the sponsor assumes that lesions within patients are independent.

The overall results were, based on 43 malignant, sensitivity 94 percent, and 159 benign, specificity 20 percent, and the confidence intervals shown there, which are fairly wide because of the small numbers here.

This is what the PPMA Amendment 5 results look like when broken down, again, by lesion type. Again, they don't add up because some categories overlap.

We see here that the sponsor interpreted these effectiveness results by lesion type to specifically include only masses. Masses gave the highest sensitivity and the highest specificity.

Of course, we see what happened to distortions. They were 100 percent on sensitivity before; now they're down to 75 percent.

Of course, calcifications were 95 percent before; now the sensitivity is 93 percent. So they've definitely shown that calcifications need to be excluded.

Now in Amendment 7 the sponsor makes an adjustment. The sponsor attempted to use Bonferroni

in response to FDA Deficiency 1(a). In other words, we specifically asked them what they were going to do about the fact that they are combining the data from the PMA Amendment 4 and the PPMA Amendment 5, and looking at all the data, how are they going to handle that?

What they said: We'll widen the sensitivity/specificity confidence interval estimates which are based on a simple, direct combination. They just added the data together, both the exploratory and the confirmatory clinical data, to test a theoretical possible set of hypotheses.

They first did the seven lesions types, which were explained by Dr. Rust: the mass, distortions, and the calcifications, and all combinations thereof, or possibly 63 lesion types, sizes and depths. Here they're multiplying seven by three sizes, and then they are breaking the depths into three sizes also.

Now these hypotheses are not explicitly included in the protocol. In the protocol they simply say: We're going to look at these things. They don't

say what they are going to do with them.

This is not statistically-acceptable because the sponsor simply estimates sensitivity, specificity, and confidence intervals for various subpopulations without actually statistically testing any hypotheses. At this point in their analysis, what they are doing is they are just telling us what the sensitivity and specificity is. I don't see any hypotheses there. Therefore, I don't see any need for any Bonferroni adjustment. Therefore, as far as I am concerned, there are no multiple comparisons requiring adjustment.

In conclusion, one thing I want to make clear is that diagnostic mammography, not just clinical examination, is required for use of BCS. The sponsor's primary effectiveness demonstration using ROC, area under the curve, loses statistical significance when mammography LOS1 through 4 is expanded by just two additional intermediate categories after excluding calcifications alone.

The sponsor's initial rejection of overall sensitivity, followed by rejection of calcification

alone sensitivity, indicates exploration which requires confirmation, which requires new data. The sponsor's attempt at Bonferroni adjustment to make sense out of putting all of the data together by widening the confidence interval estimates is not statistically-acceptable.

Thank you very much for your attention. I am going to turn the podium back over to Bill Sacks to continue with the clinical.

DR. SACKS: Before I do that, I just want to make a point about safety. We have heard that there were four adverse events out of 2,400, which is a very important aspect of safety, but there are two aspects of safety for any diagnostic device. It's not peculiar to this device.

That is the accuracy of the diagnostic output of the device also involves a question of safety. So as far as the adverse events were concerned, they were very few and minor, but from the point of view of the BCS output, we should focus on safety is more closely related to the question of sensitivity; that is, on cancers or the false negative

rate. In other words, how many cancers have their diagnoses delayed? Also, in the context of the psychological impact, the false positive rate also can be regarded as a question of safety.

Effectiveness is more closely related to the specificity; that is, its performance on the benign masses, because the intent, the intended use of the device is to save biopsies of benign masses.

Now let's look at what the clinical trial demonstrated. I am going to just summarize this briefly for the history of these again.

There were four relevant clinical submissions here: the PMA, Amendments 4, 5, and 7. After reviewing the PMA, the FDA sent a letter to the company listing a number of deficiencies, and the company's response was Amendment 4.

In Amendment 4, for their conclusions concerning the effectiveness, the company retrospectively selected from the PMA data one of two analytical indices, namely, sensitivity and specificity, as opposed to ROC curve comparison, and two of three lesion types at first, masses and

architectural distortion, and not microcalcifications.

In that same amendment, however, the revised labeling further deleted architectural distortion and referred to masses alone. So that was sort of both steps were involved in Amendment 4.

The FDA sent another deficiency letter, and the response was Amendment 5. Amendment 5 was offered as a test of the device in additional subjects. That's those 275, although not all of them were evaluable, additional subjects who had not previously been analyzed. That was because Amendment 4 had contained retrospective selections.

The company refers to this additional dataset as the "post-PMA." That is PPMA for short.

This amendment confined its analysis of the PPMA data, that is, the newly-analyzed data, just to the newly-chosen analytical index, namely, sensitivity/specificity, in the newly-chosen subgroup, masses. That was done before unvaulting that data. So as far as this data is concerned, that was prospectively done.

In addition to presenting data on a new

set of subjects, the amendment also contained an analysis, as you have seen, of the combined datasets from Amendment 4 and the PPMA.

Because of the retrospective selections in Amendment 4, the FDA asked the company to justify combining that data with the PPMA data, and the response was Amendment 7.

In Amendment 7 the company applied the Bonferroni correction, as you've heard, in an attempt to compensate for retrospective selection and the smallness of the additional PPMA sample.

Now, as we go through this, there are two overriding issues. One is the adequacy of the data, and the second is the interpretation of the data. That is, do they demonstrate safety and effectiveness of the device, assuming that we accept that the data is adequate?

On the question of the adequacy, a question that we will have the Panel consider this afternoon, and before that we will give you the questions as they are phrased more precisely. This is paraphrasing. Can the data from Amendment 4

contribute to the judgment of safety and effectiveness when it consists of retrospective selections? Is a Bonferroni correction applicable in this context? Are the data from the PPMA alone adequate for the judgment of safety and effectiveness?

In looking at the interpretation of the data, it is noteworthy for the following discussion that no formal hypotheses were explicitly put forward for testing either in the PMA or in the subsequent amendments, and let me hasten to add here that, to qualify as a testable hypothesis, there must be a quantitative criterion whereby either a point estimate may imply rejection or a confidence interval may entail exclusion.

There were two implicit hypotheses. One was that the ROC area for the device and mammography combined would exceed that of mammography alone with statistical significance.

The second was -- and this was derived from the training set of 700 subjects, of whom 150 were cancers, and the sensitivity of the device -- that is, the threshold for the device was set such

that 149 of those 150 cancers were positive, with one being negative, which is a sensitivity setting of 99.3 percent. There was an implicit hypothesis that the point estimate for sensitivity would be at least 99.3 percent in at least 75 percent of simulations with the data.

The protocol otherwise contained only non-quantitative statements of what the company hoped to achieve in the clinical trial. One example, quote, "The objective of the study is to determine if the BCS, when used in conjunction with clinical examination and/or diagnostic mammography, increases the ability of physicians to differentiate benign from malignant or suspicious breast abnormalities." But there is no quantitative criterion by which we can judge success or failure on this, except through ROC area comparisons, but those were dropped.

In the original PMA submission, the comparison of ROC areas failed to achieve statistical significance except, as Harry has shown you, as an artifact of too few points in the mammography alone curve. It was, therefore, not pursued in any of the

amendments.

In addition, the sensitivities failed to achieve a level of 99.3 with 75 percent confidence in any of the datasets. Here is a diagram that shows them. I finally get to use this pointer.

This is the upper righthand corner of an ROC plot. It is about a quarter in both dimensions. Mammography alone, because of this being the universe here is women on their way to biopsy based on mammography, was 100 percent sensitive. That is just an artifact of the choice of the universe here.

It was, similarly, zero percent specific. That is, there were no non-biopsied people here.

Now the PPMA -- I'm sorry, the original PMA point estimate with 187 cancers -- this "N" is just the number of cancers -- turned out to be 97.1 percent. For reference, this line here is the 99.3 percent level that was involved in that implicit hypothesis of trying to keep it above 149 out of 150. Its confidence limits are, as you see here, 94.1 to 98.8.

The next data was Amendment 4, where out

of this set of 187 were culled the 90 masses. It is the same as -- these 90 are part of this on the 97. In those 90 we got point estimate of 100 percent sensitivity. The higher confidence bound, of course, is also 100 percent because you can't go over that, and the lower one is about 96.7.

Because of the retrospective selection of these out of this group, the next set of data in Amendment 5 was the PPMA, of which there were 15 cancers. The point estimate there was 93.3 percent because one of those turned out to be negative, so 14 out of 15.

The confidence interval on this, because the number is so small, 15, is rather wide. Interestingly, the lower confidence bound is actually below the chance line.

When the two sets of data are combined, if you think it is valid to combine these two, you get a point estimate of 99.0, which is still below the 99.3, and its lower confidence limit is about 95.6. So that sort of displays all of the data in reference to that 99.3 implicit hypothesis.

The potential safety and effectiveness in the U.S. population as a whole -- this is a bit of a busy slide, but I'll walk you through it. The percent of U.S. biopsies that are potentially obviated by the BCS, if used on all eligible women, and we've seen these figures, and mine are very close to the company's, 1.3 million U.S. women biopsied each year, of which I use the number 45 percent; the company used 45.5.

Forty-five is not only a typical figure for the country at large, but happened to be exactly the percentage in the used data combining the PMA and the PPMA; 45 percent of them were cancer. So I used that figure. That's about 585, which is very close to -- Steve Rust gave you a figure that was 591,000, very close -- of which 80 percent, roughly, are benign. That's about 468,000, of which 15 to 20 percent, using the various ranges of specificity that we got for the device, which would be 70,000 to 94,000, would be BCS-negative and, therefore, save the biopsy.

So 70,000 to 94,000 out of 1.3 million is roughly 5 to 7 percent of the 1.3 million U.S.

biopsies would be obviated. That is if the BCS were used on all 585,000 women who are eligible; namely, mammographic masses on their way to biopsy.

In addition to saving biopsies on these benign masses, approximately 1 to 6 percent of the malignant masses -- that's, again, the range from the data -- and a half to 3 percent of all breast cancers, that is, not just of masses, might be delayed in diagnosis.

A couple of labeling issues involved the size of the mass and the depth of the mass. We are going to ask you for some discussion on this this afternoon.

The size of the mass: The effect of small lesion size on device sensitivity was difficult to evaluate since only 2 out of the 105 cancers in the two combined sets were smaller than 5 millimeters. Here are the figures for how they fell. This is different from the figures that Karleen Callahan gave you, but she was including the ones that we didn't have the data for. This is just the two combined, Amendment 4 and PPMA data. Only two of the malignant

masses were less than a half a centimeter, 5 millimeters. So it is hard to make any statement about it.

With the chosen threshold, there was no definite effect of lesion depth on BCS result, but, as the mammographers here know, the effect of lesion depth is difficult to evaluate because depth is not easily gauged on the mammogram. Worse yet, we are imaging women in a position in which the breast is pendulent. It is a fairly mobile structure. Depth is variable. A given lesion has different depths in the breast, depending on the position.

Therefore, we really have difficulty from this data making any judgments or conclusions about depth. However, one should realize that, just from the physics of the situation and the physiology, that the deeper the lesion, the less effect a cancer will have on contrast of temperature on the overlying skin. So that might affect device sensitivity, but we can't make any statement about it.

Conclusions then: In summary, only 4 out of 2,407 subjects had an adverse event, all minor. In

that regard, the device seems safe.

There were no explicit, quantitative hypotheses. There were two implicit, quantitative hypotheses. Neither hypothesis was fulfilled. Most of the data was selected retrospectively. Bonferroni correction we feel is not applicable in this context, in part because there were no hypotheses; therefore, no alpha levels to protect, and so on. But if you did use the correction to widen the confidence limits with the point estimates already below the implicit hypothesis of 99.3, that doesn't help keep them above it.

Finally, using the trial results, if the BCS were in general use in the U.S., it would obviate 5 to 7 percent of the 1.3 million biopsies a year, and approximately 1 to 6 percent of these obviated biopsies would turn out to be malignant and their diagnoses would, thus, be delayed.

The people who are still awake will notice this 1 to 6 percent is not the same figure as 1 to 6 percent I gave before because one was looking at the percent of malignants that would be negative and this

is looking at the percent of negatives that would be malignant. It turns out that the diagonal members of the 2x2 table are about the same, so these figures come out to be the same, or perhaps it's not so coincidental.

Thank you.

CHAIRMAN MEHTA: I think, before we leave for lunch, I would like to remind you that the open Committee deliberations will resume at 1:00 p.m., but the Panel members are requested to be back here at 12:30 for a Panel-members-only Closed Session from 12:30 to 1:00 p.m.

(Whereupon, the foregoing matter went off the record for lunch at 12:00 noon and went back on the record in Closed Session at 12:36 p.m.)

A-F-T-E-R-N-O-O-N S-E-S-S-I-O-N

1:03 p.m.

CHAIRMAN MEHTA: Before I call the meeting back to order, Nancy Brogdon, Director of the Division of Reproductive, Abdominal, and Radiological Devices of the Office of Device Evaluation, has a few words that she would like to say.

MS. BROGDON: Thank you, Dr. Mehta.

I wanted to let you know that two members of this Panel are completing their terms before the next scheduled meeting.

First is Ms. Marilyn Peters, the Panel's consumer representative. Ms. Peters was a Patient Health Education Coordinator for the Department of Veterans' Affairs at the West Los Angeles Health Care Center. She recently retired from this position.

Serving as the consumer representative is a difficult task, as you know, due to the wide spectrum of devices that come under the radiology umbrella. Although meetings of the Panel required Ms. Peters to travel across the country, her attendance has been perfect during her tenure.

Thank you, Ms. Peters, for your service on the Panel.

(Applause.)

Next is Dr. Alicia Toledano, the Panel's biostatistician. Dr. Toledano is Assistant Professor in the Center of Statistical Sciences at Brown University.

Dr. Toledano has served the Panel well, providing insightful input on all the various devices that have come before the Panel during her tenure. In addition, she's been called by at least one other panel for her statistical expertise and, when requested, she has provided written reviews that are very comprehensive for statistical analyses of various PMAs. We'll miss her contributions as a Panel member, but, if she agrees, we hope that she will stay on as a consultant, so that we can call on her for other difficult statistical issues in the future.

And thank you, Dr. Toledano, for your service on the Panel.

(Applause.)

We'll be sending both of you recognition

plaques as soon as they're signed by our new Commissioner, Dr. Mark McClellan. Thank you very much, and best wishes to both of you.

CHAIRMAN MEHTA: I would now like to call the meeting back to order.

I would like to remind the public observers at the meeting that, while this portion of the meeting is open to public observation, public attendees may not participate unless specifically requested to do so by the Chair.

We will now continue with the Panel's discussion of the PMA. What we will do here is I would like to ask the Panel to pull out the Panel discussion questions, which were actually placed in your blue folders, and maybe use these as a starting for discussion points for the Panel discussion.

Who would like to go first?

DR. TOLEDANO: Are you sure you want me to go first?

CHAIRMAN MEHTA: Yes, if you would like to, go ahead.

(Laughter.)

DR. TOLEDANO: Okay. My first question for today has to do with the feasibility of using this device in clinical practice. We have seen numerous exclusions, many of them for technical factors. We have reference to the fact that the device has changed, more things are automated, there's better training, the company thinks that the performance in terms of being able to obtain results on women has improved. So I guess my concern is, have you proved to me that you would be able to obtain results in the women who undergo this procedure?

Now there's a second thing that's sort of buried in some of the appendices, and maybe it comes to my mind easily because I'm a statistician. The blinded investigators were assigned three to each subject. If all three -- when it happened in the first study, the subject was reviewed by two investigators, and if there was a disagreement, the third investigator was brought in.

You can explain if I've got this in any way incorrect. For each woman there were up to three scores weighted one-third each, but there were some

women who only had two evaluations and there were some women who you could only get one evaluation and there was some women who you got no evaluations.

So even beyond everybody who was excluded because you got no evaluations, there were instances where you tried to get evaluations but you couldn't. That, to me, is a big concern. So could you address that a little bit?

CHAIRMAN MEHTA: Does the sponsor want to suggest someone who might be able to address that?

DR. RUST: Let me start just by saying that I think you do have it correct in terms of how we analyzed the data, but, just for everyone's benefit, we attempted to get three evaluations for each lesion. If we got three, all three were used in the analysis. We weighted each evaluation one-third. In some cases there was a failure to evaluate on a single, leaving us with two evaluations. Those were both included, weighted one-half. In some cases we failed to get two but got one, and that was given a weight of one in the analysis, and, yes, there were cases where we failed in all three cases to get an evaluation. So I believe

that that is the correct interpretation of how we handled the data.

Lynn Satterthwaite will address the feasibility of using the device.

MR. SATTERTHWAITE: We'll address it in two parts, the device itself and then the other part of it Dr. Callahan will address.

In the clinical trial we utilized a device that, as we have characterized earlier, was mostly a manual device. Manual had much to do with the cooling challenge itself, but there's a little more to it than that in that the images were checked well after the fact, well after the imaging session.

To address several of the problems, problems included movement of the patient, poor positioning of the patient, and the possibility that the cooling challenge wasn't adequate. Much of that had to do with the reliance on the technologist to manually throw a switch. The protocol required that they watch the second hand, and 30 seconds into the imaging procedure they were required to turn that switch. In many cases that switch was not thrown, it

was thrown right at the beginning, or thrown too late, and those kinds of problems.

So these issues have been addressed in the following way: The automated system, the functions performed by the computer now include doing this temperature challenge check right at the imaging session time. That will prevent us from accepting that image if it was not adequate.

That takes into consideration all factors. Was it adequate? Was its timing done in the right place? It also prompts the technologist in the session to review the image at that point in time. So there's some discipline here that we enforce because the technologist must look at the screen and go through it step by step.

They look at a cinema view, much like you saw today, and that would allow them to make the decision as to whether they truly were positioned correctly and whether there were no artifacts. At times we would see a gown strap that's down in the image and those sorts of things, and those cause problems.

So we believe that the design of the system, which functionally does exactly the same thing, the same camera, the same cooling -- it is just the protocol is enforced by the computer -- will solve many of these problems. We do believe that in clinical practice we can enroll patients, at least from an imaging standpoint, and get good images or know that we don't have good images and we'll know to do it again, if we have to.

DR. CALLAHAN: And I'll just briefly address technologist training. Another thing that we have instituted since identifying that there were a number of images that were not usable is we have developed a set of training cases to use for technologist training. We've implemented that in our study with Dr. Hughes at Mass. General. That study just began a few weeks ago, but I understand the first eight or ten patients that have been enrolled, the images have been taken; there's been no quality problems. They are evaluable images, albeit a small number.

MR. SATTERTHWAITE: And let me follow that

up. When we enroll patients in the clinical trial, if any of the factors that made their image inevaluable were there, then we would lose that patient. With the new, automated system and the prompting by the computer, we would know to go back and do the imaging then. So there is a number that we wouldn't lose because of that factor.

DR. TOLEDANO: Okay. So can I follow up?

CHAIRMAN MEHTA: Yes, go ahead, Alicia.

DR. TOLEDANO: Okay. So that addresses the issue of being able to obtain usable images from the system. Now the physician has to look at them and obtain a score, a test result.

We are still left with the issue that in the trial images that were potentially evaluable that were presented to physicians, they couldn't obtain a test result. How often? What's been done to address that?

DR. HUGHES: What we have been dealing with is that the central reviewers received a set of mammograms and then had to figure out where the lesion was relative to the mammogram. Any lesion that didn't

show up on a mammogram they could not use.

Within the study we're currently doing is, how do we incorporate it into clinical practice? The way we're doing this is I'll examine a patient. I'll review the mammogram. I'll schedule that patient for a biopsy. We'll then do the imaging. I know exactly where the lesion is. I know where I felt it. I know where it is on the mammogram, and I can pick where to go.

Is that your question?

DR. TOLEDANO: That's not the question.

DR. HUGHES: Then what is your question?

DR. TOLEDANO: Okay. The question is, three physicians were presented with images that they were supposed to be able to work with. Sometimes not all three got it. How often? How do you address that in clinical practice?

DR. HUGHES: Within clinical practice I can hit the area of interest for the patient that I'm dealing with because I know where the region of interest is. Other than that, I'm not sure what the three disagreements are.

DR. TOLEDANO: I'm not stating it clearly?

DR. CALLAHAN: The question, if I understand it correctly, is we had three evaluators. They each had the same films. They each had the same thermal image. They each were given the same opportunity to evaluate, and in some cases not all three chose to evaluate the case or were not able to evaluate the case.

The issues are partly as Dr. Hughes described, in that they may not have felt -- there's differences in mammographers' interpretation -- they may not have felt, one of the three may have said -- well, if you've seen the evaluation sheet that the evaluators got, it described the case to the extent that was in the patient database as to where the lesion was located, as well as the lesion descriptor. So it says there was a mass in the upper outer quadrant of one centimeter size, for example.

If the evaluator felt like he or she did not see that lesion with that descriptor, they would choose not to evaluate it because there could be other things on that mammogram perhaps that they thought was

suspicious, but they were instructed -- the instruction on the evaluation sheet was to attempt to evaluate the lesion that was actually biopsied.

So that accounted for some of the cases where not all three evaluators interpreted the information in the same way. I think that was the predominant situation.

CHAIRMAN MEHTA: Let me follow up on that question. The use of the expert panel for evaluation was carried out through an amendment. It was not the original intent of the protocol to have such a panel.

DR. CALLAHAN: That's correct.

CHAIRMAN MEHTA: Why was it done? When was it done, and how many patients had already been enrolled and evaluated prior to the panel coming into place?

DR. CALLAHAN: Okay, that amendment was done in September of 1998. My understanding is that that was done for a couple of reasons.

One is it was felt that the original investigator might, by knowing the patient outcome, biopsy outcome, or something, their evaluation might

potentially be biased at that time, if their evaluation of the IR image could have been, you know, biased by the pathology outcome.

CHAIRMAN MEHTA: But the original investigator did not have the pathology at the time of the IR image.

DR. CALLAHAN: There was no requirement -- and Dr. Parisky can perhaps address this or Dr. Hughes -- but there was no requirement. They did the IR imaging prior to biopsy, but they were not required to do the evaluation in real time at the clinical sites.

CHAIRMAN MEHTA: But they did the ROI in real time?

DR. CALLAHAN: No. No, they did not.

DR. PARISKY: I'll demonstrate a patient flow. I would enroll a patient, get informed consent for the IR imaging and also informed consent for biopsy, and would take the patient because I have a mass. Rather than do an ultrasound, see the mass, take the patient off the table, take them to another room, do a test, and then return for the biopsy, at my center specifically Dr. Silverstein prides himself

that we will give you a result within that day.

So I enrolled the patient, would go to ultrasound, would do a biopsy, and while I can't prove to you scientifically, those of us who do that, the way we watch a needle go into a lesion, we can tell you if it's likely cancer or not. We felt, in discussion amongst ourselves, or we had information such as an MRI or we had information such as obtaining prior films, that showed a rapid growth, things that would be biasing the investigators at that site that perhaps would have a bias.

It was felt by the company and in counsel with the principal investigators at the various sites that, to try to remove this non-quantifiable bias, that it would be evaluated best by a separate panel.

CHAIRMAN MEHTA: Let me follow up on that. Given the fact that the assessment is actually done by a computer using a number, it's either a yes or no, all it would have taken would have been for the diagnostic radiologist to place the ROI, and the computer could have done it at any point because the computer didn't know what the pathology was.

DR. PARISKY: If this was all in a work area that was immediate within my clinic or within some of the clinics, I think, and if patient flow allowed for that. While hindsight may have suggested that, the course of events were that we would in a busy clinical practice move a patient from the IR table, you know, pull together the images necessary for the biopsy -- this is my situation -- and proceed forth.

I appreciate your comments and criticism.

CHAIRMAN MEHTA: So in that case, in your busy clinical practice, how do you see incorporating this and giving the patient the results on the same day, if you can't do it in real time?

DR. PARISKY: The same way I incorporate giving them my results to each of my patients following diagnostic workup. I will incorporate that as -- will then take the time to do so, once all the information has been compiled.

Conducting research is one thing. Providing patient care adequately, we'll make time for it, sir.

DR. CONANT: How long does it take to get the IR reading?

DR. PARISKY: It takes minutes, but in terms of locality as to where the patient was taken, in some instances it was not on the same floor.

DR. CONANT: No, the numerical IR output?

DR. PARISKY: Three seconds, yes.

CHAIRMAN MEHTA: Let me just follow up, and anyone can answer this question. The reason I'm sort of going on this direction of questions is, to some extent, the use of a panel is useful. To another extent, it detracts from reality in terms of how this machine is intended to be used. It is intended to be used by a practice in reality, not necessarily at the Norris Cancer Center, where there are hundreds, if not thousands, of women that are undergoing mammography on a daily basis, but much smaller numbers of women that are having mammography.

The ability in a clinical trial to demonstrate the effectiveness of this machine in that context is negated by using a panel that replaces an individual radiologist's decisionmaking. That is one

of my concerns.

DR. HUGHES: Just from a logistical point of view, I have the machine in my clinic. I'll see a patient. I'll say, "You're going to need a biopsy." I'll have her consented for the biopsy. She will go and have her thermography done. It takes about 25-30 minutes to do.

The image is immediately available for me to look at, and I evaluate the image immediately at that time. I'm not using these images to change my care, because that's not part of the protocol, but getting the image done, seeing it, putting the ROI in place, getting a reading out is done within the context of my normal clinic day.

CHAIRMAN MEHTA: And I'm sorry, do you do this in the context of a clinical trial right now or what?

DR. HUGHES: This is a clinical trial we're running currently, which is --

CHAIRMAN MEHTA: It's a separate clinical trial from --

DR. HUGHES: That's correct, which is to

look at -- I guess, to answer your question, and we want to answer the same question: When we put this into clinical context, how much does it help us with our patients? How logistically is it possible? How often are we not able to find an ROI? That's not happened as of yet. How often are the images unevaluable? That's not happened as of yet.

And this is eight or ten patients, but we're not having the problems that you have when you're taking these images, shipping them across the country, and asking somebody else to evaluate them.

DR. CONANT: So these are prospectively patients that are coming to your office after having their mammogram and ultrasound?

DR. HUGHES: These are patients who are in my office who require a biopsy. They may or may not have had a mammogram or ultrasound, depending on their age. But at the end of the visit, when I say, "You need to have a biopsy done," based on mammography, ultrasound, whatever is available, my exam, we then do the imaging and use that to predict what we would do, were we to use this as an FDA-approved device, which

it currently is not.

DR. CONANT: So they are having mammography and ultrasound before you recommend the biopsy?

DR. HUGHES: Well, if the patient is 20 years old, I don't do mammography.

DR. CONANT: No, but ultrasounds.

DR. HUGHES: So in the majority of cases they would have a mammogram, ultrasound, and physical exam, that's correct.

DR. CONANT: And how many of them go to biopsy that have a negative ultrasound and a negative mammogram?

DR. HUGHES: Out of ten patients, I can't answer that quite yet. But within my practice I have a fair number of women who have come in with breast lumps that turn out to be fibrocystic changes that the patient is concerned about. That may be 10-20 percent of the biopsies I see.

DR. CONANT: With negative ultrasounds?

DR. HUGHES: Negative ultrasounds, patients with calcifications, patients with

mammography, and this is a different trial.

But I can't give you the percentages of those things currently.

DR. CONANT: Okay. Thank you.

CHAIRMAN MEHTA: Alicia, go ahead.

DR. TOLEDANO: So going back to the patient flow, they all are recommended for biopsy. So you've done all your preliminary workup. They're all recommended for biopsy. You do the infrared imaging. They proceed to biopsy.

Now in the clinical trial results that you present to us in support of your PMA you have excluded the women who did not have biopsies, even, for example, if they went for the biopsy for like an ultrasound-guided biopsy, and the ultrasound determined that it was a fluid-filled cyst.

Now you're looking at two different populations, because I cannot tell, when I'm doing the IR imaging in clinical flow, I can't tell whether they're going to end up having a fluid-filled cyst. I don't know if it is unguided biopsy.

Doesn't the exclusion of, I think it was

20 percent of women, 10 or 20 percent of women, because they didn't have biopsies, doesn't that exclusion open up the results of your clinical trial to some significant bias?

DR. HUGHES: I'll let the statistician handle that one, I think.

(Laughter.)

DR. RUST: I think asking the question whether or not those exclusions create bias is a very legitimate question. Now one of the significant classes of patients that were excluded that way are patients that, upon ultrasound, would not have been scheduled for biopsy.

So I think what we have here is a sequencing-of-events type of thing. In Dr. Hughes' clinic that ultrasound occurs before he makes the decision on IR imaging. In our case, because of the enrollment criterion for the trial, we enrolled based on mammographic findings and clinical findings only. So a patient got enrolled and had the ultrasound post.

But I believe that what you end up with are the same populations in both cases, that the

population we end up studying, analyzing, is the same population that Dr. Hughes ends up sending to the IR procedure. So, in fact, those exclusions I believe are a good exclusion as opposed to a possibly negative exclusion.

DR. CONANT: I think Dr. Callahan this morning mentioned, in going over the exclusion groups, that some of them were complex cysts or questionable cysts that on aspiration, or were aspiratable but a needle was placed in them. I think they probably had already a recommendation for intervention, though, which would be a biopsy recommendation, and then they were aspirated.

I think that's a subgroup that Alicia may be referring to that actually is suspicious on mammography, whether they have the ultrasound or not, where complex or maybe even appearing solid -- this happens quite a lot clinically -- then had an aspiration, then were excluded. That's a bias, I believe, on those recommended for biopsy, excluded before they got to the analysis.

DR. PARISKY: The protocol that was

reviewed with the FDA required histological proof of disease. While the scenario that you address, you're absolutely right, that needle was stuck in there, but any specimen that you might get from an intervention like that would be cytological and also fraught with the known cytological problems. Strictly following the approved protocol, we were required to disenroll patients who did not have histology.

DR. CONANT: Did you then have patients that might have had an FNA in someone's office for a suspicious area on exams that were positive and malignant and, therefore, known to the enrolling radiologist, an FNA based on palpation that wasn't histologic but --

DR. PARISKY: In my own center there was one case that was excluded because we knew the answer. We voluntarily excluded it on an ethical basis, but it didn't fit the criteria.

DR. CONANT: Okay, but these women still would have had IR imaging in the flow --

DR. PARISKY: I don't remember if they went to completion. You know, we enrolled patients,

and then some decision points were made actually at times before IR imaging, you know, because we would get a result back. Sometimes patients were -- yes?

DR. CONANT: I'm just trying to think how it fits into the clinical flow of things.

DR. PARISKY: Yes. I think the way Dr. Hughes -- in terms of a standard clinical flow, I would give my own experience, on mammographically-apparent lesions we would run parallel or complementary to ultrasound. In a negative ultrasound we would then perform the IR. If it was a solid mass, we would perform the IR. If it was a cyst, we wouldn't perform the IR in my clinical practice, and it would be done within the setting of the two hours that the patient's in my hands.

DR. CONANT: There's been some good data published recently about palpable areas, which I suspect you're including in this flow, palpable areas, negative mammogram, negative ultrasound, and negative predictive value. It's almost 100 percent. Dan Kopans, Mary Scott Soo, Sue Weinstein, they've all -- three different publications.

So I'm just wondering how this would fit in and influenced --

DR. PARISKY: I would welcome the opportunity to evaluate patients with palpable abnormalities in a setting of a negative or a dense mammogram usually setting, and it would be complementary to ultrasound.

The problem we had in generating numbers was that patients who actually fit that scenario, not all centers placed markers, metallic markers, on the mammograms to allow the tertiary parties to --

DR. CONANT: Okay. So those would have been excluded? Those were part of that exclusion? Okay. I see.

DR. CALLAHAN: I think that both Dr. Toledano's and your questions are insightful. I think we just don't have all the data that you would like to know about how, you know, the impact or the potential impact of these other diagnostic procedures.

That is one of the points of the protocol at Mass. General, is to look at how this will integrate into actual clinical practice. There the

enrollment criteria is women, mammography and/or ultrasound, whatever diagnostic tests lead to the decision for biopsy.

DR. TOLEDANO: Could you take your existing data and re-analyze it in an exploratory manner as if you had separated out masses in the beginning, had the ultrasounds before the IR? Could you do this? And if you could do it, have you done it?

DR. CALLAHAN: We could take our existing data and we certainly have a case report form for all those that have mammograms. I mean that was part of the information that was collected, as well as those that were enrolled only on clinical examination; that is, there wasn't mammographic data available. We could do an analysis like that.

Does that answer your question?

DR. TOLEDANO: Uh-hum.

DR. RUST: I think we can do the separation based on mass/non-mass in the beginning, as you're indicating. We do not have data on why the biopsy was not performed. So that's where our

inability to do that analysis --

DR. CALLAHAN: And let me make it clear, we obviously couldn't do -- I was talking about potential bias in patient sets between those that weren't evaluated.

DR. TOLEDANO: That's right.

DR. CALLAHAN: Obviously, if they weren't evaluated, we don't have the IR data --

DR. TOLEDANO: Right.

DR. CONANT: -- so we could do bias, you know, potential bias.

DR. TOLEDANO: Dr. Brenna and I had the same reaction to those answers.

DR. GENANT: Well, overall, I have concerns about the primary efficacy and just how robust the examination is. I think that we do see that the primary efficacy endpoints of the area under the curve did not reach levels of significance and hardly showed even a trend.

Then it was in the post hoc analyses where you did reach levels of significance of sensitivity/specificity, but I do believe that the

PPMA data cannot stand alone, and the Amendment 4 data are basically post hoc. So I think that you have problems with regard to statistically-significant endpoints in this study.

I think on top of that, we have been discussing the relevance of the manner in which the clinical trial was conducted to the way in which you anticipate using the system clinically. Now this is often a problem when you're doing clinical trials and then you're trying to extrapolate to the clinical practice, and I recognize that.

But I think it is also clear that there are many issues that at this stage could be better addressed in a prospective study than, in fact, they were, as we look back on the way that this study was conducted.

I remain particularly concerned on the evaluation side of the images. I don't really have a sense of how robust that is. You've told me that, for example, among readers, between readers, that one would see perhaps plus or minus four points; that on, say, a 100-point scale, that doesn't sound like it's

terribly critical.

On the other hand, since you are using a threshold-based yes/no, it would be particularly important to know in the cases that are perhaps plus or minus ten around the threshold what kind of reproducibility there was, in fact, from one reader to the next.

And do you, in fact, have to have a trained panel of three readers reaching a consensus for this type of an approach to work or can a single reader, relatively reliably, give consistent information? I don't know that from the data that you have shown us to date. I think that is extremely critical when it comes down to using this in practice.

DR. RUST: Would you like me to respond?

CHAIRMAN MEHTA: If you would like to.

DR. RUST: I guess the last one first, and then I might have to ask you to remind me of the earlier points.

In terms of the way we analyzed the data, we allowed ourselves to be penalized for reader-to-reader variability within a lesion. In other words,

we did not take the three evaluators, reach a consensus evaluation, and then evaluate that against the gold standard.

We did, in fact, if a lesion went two-thirds one way and one-third the other in terms of its evaluation, then that two-thirds error versus one-third correct is included in our performance statistics. So we did penalize ourselves for reader-to-reader variability within lesion in terms of positives and negatives. Okay?

Go ahead.

DR. GENANT: Well, by virtue of the way that you did it, I'm not certain that actually three readers, three trained readers working kind of in concert -- in a sense, you're using their data combined -- may be more reliable than the single reader would be.

We simply don't have data, you haven't shown us data, on the impact that that might have. Yet, it has a lot of relevance about how you could, in fact, use this in practice.

DR. RUST: Yes, I guess our rationale --

and it's easier to think, if we have two evaluators -- our rationale was, if you have two evaluators, and let's say it's a malignant patient. One calls it positive; the other calls it negative. We would consider that a lesion for which it's more difficult to determine whether it's positive or negative, and so we put part of that lesion in the correctly-classified category and part of that lesion in the incorrectly-classified category.

So we are, as I think you can see by that explanation, penalizing ourselves for having multiple readers, and them not necessarily agreeing. So there was no averaging through consensus.

DR. GENANT: Well, I think that one could look at this and simply determine what the threshold was for each of the three readers and how that was in terms of positive/negative, and see what that level of agreement is. We don't know, but I presume that you probably have done that analysis at some point, I would think.

DR. RUST: Looking at the threshold for each reader?

DR. GENANT: Yes, and then also a reader doing the same examination twice in terms of, do they place the region sufficiently nearby that it doesn't impact your result? I don't know that.

DR. RUST: We did not have the same reader place the same ROI twice because we felt we couldn't get an independent placement the second time. Our reasoning was that, if you have three readers doing it, that includes both within-reader variability and reader-to-reader variability in your observed variability. That was our rationale.

But you had other points earlier. I don't know if you want me to address any of those.

DR. GENANT: Well, I was basically commenting on, I think, some of the deficiencies with regard to how robust the statistical analyses were and what we can draw from that. Because I simply feel that, based upon the information that we have, that I don't have a high level of confidence that the device is effective.

DR. RUST: Okay. You specifically indicated that our analysis of the original study

data, you considered that to be post hoc. I guess what I would like to ask the Panel to consider is whether or not that analysis is truly and entirely a post hoc analysis.

What I have to draw your attention to is that in the protocol there was a prospectively-placed plan to do analyses by lesion type. Given that that was prospectively planned for, that bounded any of the analyses that would fit within that plan.

We specifically have corrected our analyses to be correct for any of those analyses in that bounded plan. Therefore, in my judgment as a statistician, it's not subject to being a post hoc analysis because it was done according to a prospectively-planned set of analyses.

There is a necessity to correct significance levels associated with the inferences, and we feel that we have done that. So I would just ask you to reconsider your thoughts on whether it is post hoc or not.

CHAIRMAN MEHTA: Dr. Rust, as you point us to the direction of looking at the protocol, to look

at the plan, could you also point out to us where in the protocol the quantifiable variables for specificity and sensitivity, for analysis by mass, are included in the statistical evaluation plan in the protocol?

DR. RUST: I'm going to, as far as hypotheses and objectives, I'm going to turn to Dr. Callahan.

DR. CALLAHAN: You're correct, there is no quantitative definition in the protocol.

CHAIRMAN MEHTA: Thank you.

Alicia?

DR. TOLEDANO: There's a lovely article in AJR which you have an article coming out in January. It is in 1996 by a woman named Nancy Obuchowski. Dr. Obuchowski is one of our foremost people in statistics and diagnostic radiology.

In that article she discusses conduct of multi-reader studies and proper analysis of multi-reader studies. The convention in this field, in the field of multi-reader studies, is to present the measures of diagnostic accuracy per reader and then

overall.

I do not see anywhere, and I have looked through everything, sensitivity for masses for reader one in your post-PMA or even within your pre-PMA dataset. Reader one, sensitivity, specificity, AUC; reader two, sensitivity, specificity, average. I don't see that. I need that. Do you have that?

DR. RUST: In fact, we did do such an analysis in the original submission, and there is a Table 5.9-M --

DR. TOLEDANO: Tell me which one is that. Module 5?

DR. RUST: Module 5, page 505.

DR. TOLEDANO: I still don't think that's what I need.

DR. RUST: Well, I don't think it's what you need, either, because it goes all the way back to the original submission, but I did, in answering your question, I did want to point out that we, in fact, did that analysis in the original submission, and you are correct, we have not repeated that analysis in any of the subsequent submissions.

DR. TOLEDANO: Okay.

CHAIRMAN MEHTA: Dr. Hooley, do you have any questions?

DR. HOOLEY: Yes. I'm concerned about the broad definition that you use for masses. In clinical practice a mass can have various definitions, meaning that one can be a mass as only seen clinically. Another one is the mass is only detected on mammography. The third definition is a mass that is detected on mammography and ultrasound as well.

I feel that ultrasound has been neglected in your review of the decisionmaking process. I am wondering if you have any data on the ultrasound characteristics of the masses and if that is available.

DR. CALLAHAN: The short answer to that question is we did not collect -- it was not part of the protocol, and we did not collect ultrasound data. So it's not available.

DR. HOOLEY: But, yet, in clinical practice almost all of these masses will undergo ultrasound, and ultrasound can be used to determine